Stereo Vision

Richard Daelan Roosa

Summer 2020

1 Introduction

This document seeks to explore the effects of the physical parameters of the stereo camera system, and the physical limitations they impose on any depth-determination algorithm applied to the same. The most basic stereoscopic camera array (like the ones studied here) consist of a pair of cameras aligned horizontally with a well known horizontal dispacement. They are aimed such that the central axes of their fields of view are parallel. This creates a region where the two cameras' respective fields of view overlap; objects in this region appear images produced by both cameras. Such objects will appear displaced horizontally between the two images. For example, an object appearing in the center of the image from the camera on the left will appear somewhere to the left of the center of the imgae produced by the camera on the right. The extent of this "binocular disparity" will be dependent on the distance the object is from the camera array, and can therefore be used to calculate the distance from the camera array to different features common to both images.

2 Physical Parameters of a Stereoscopic Camera Array

There are a few important direct physical characteristics of the cameras themselves that impinge heavily on the practical application of such an array. They are labeled and will be hereafter referred to as follows (see figure **??**):

- Horizontal and vertical resolution (r_x and r_y): the width and height of the image (in pixels).

- Horizontal and vertical field of view (Φ_x and Φ_y): the angular width and height of the image (in degrees).

- Image central axes ($\vec{C_R}$ and $\vec{C_L}$): the axis in each camera bisecting both Φ_x and Φ_y. $\vec{C_R}$ and $\vec{C_L}$ should be parallel.

- Horizontal displacement (l): perpendicular distance between $\vec{C_R}$ and $\vec{C_L}$.
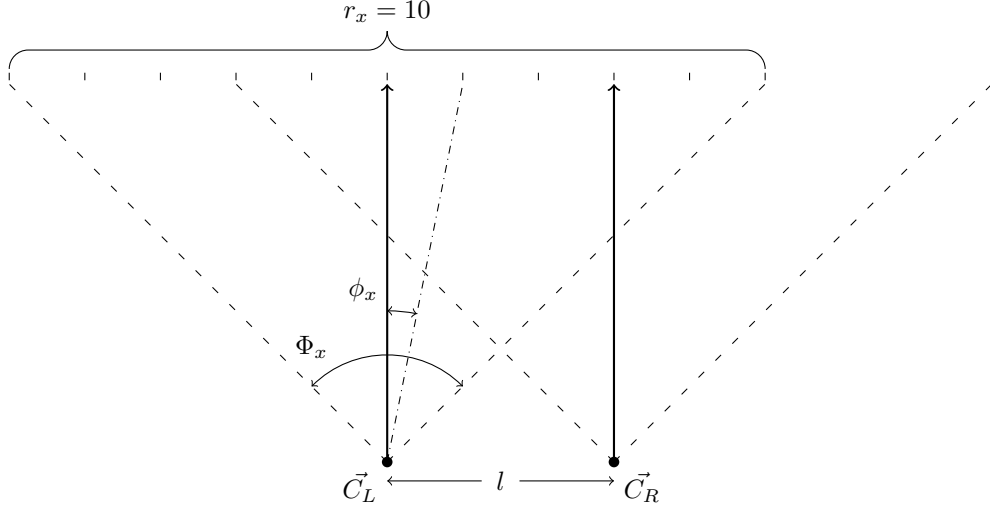
Figure 1: Top-down diagram of a stereoscopic camera array.

Camera images are generally presented in as two-dimensional rectangles (for example, on a sheet of 2D paper or a 2D phone/computer screen) and are generally thought of consisting of rectangular pixels. However, in reality, a camera image is two-dimensional representation of a spherical rectangle,[1] of which each pixel is a section, itself a spherical rectangle. Therefore, the width of a pixel can be used as a unit of angular measurement, where each pixel represents an equal proportion of the total field of view. The angular width of a pixel in the horizontal dimension (ϕ_x) is given by its proportion of the total horizontal field of view of the camera.

$$\phi_x = \frac{\Phi_x}{r_x} \tag{1}$$

Therefore, for a standard definition video camera (with a resolution of 640x480 pixels) and a 45° horizontal field of view, each pixel represents 0.07° degrees. An object appearing 100 pixels to the left of the center of the frame lies on a line passing through the camera's focal point, 7° displaced from the central axis of the image. These types of measurements form the basis of the depth determination algorithm.

---

[1]A spherical rectangle is a region of the surface of a sphere bounded by two pairs of planes, where each plane passes through the center of the sphere, and the lines of intersection formed by each pair are perpendicular.
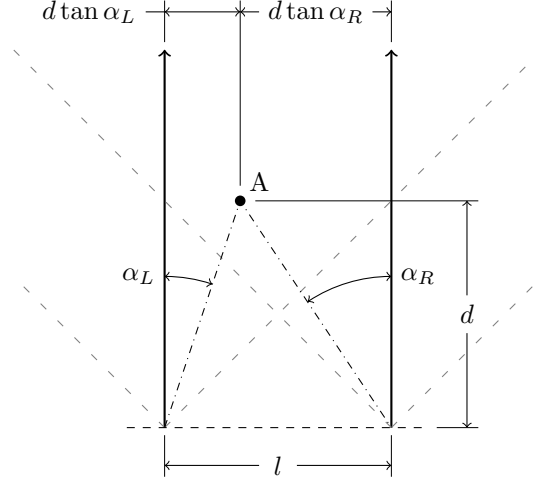
Figure 2: An oject in stereoscopic view.

3   Depth Determination by Binocular Disparity

Consider figure 3. Object A is positioned such that it is within both cameras' fields of view; it will appear in both images. Also, as described above, based on the pixels in which Object A appears in both images, we can determine α_L and α_R From there, we can determine trigonometrically that

$$l = d(\tan \alpha_R + \tan \alpha_L)$$

$$d = \frac{l}{\tan \alpha_R + \tan \alpha_L} \tag{2}$$

However, in that example, the object falls within the region between $\vec{C}_L$ and $\vec{C}_R$. In figure 3, consider the case where Object A falls outside and to the left of that region, but still within the region of overlapping fields of view. Now, we see that trigonometrically;

$$l = d(\tan \alpha_R - \tan \alpha_L)$$

$$d = \frac{l}{\tan \alpha_R - \tan \alpha_L} \tag{3}$$

Now consider if Object A were placed on the opposide side of the figure. It would yield

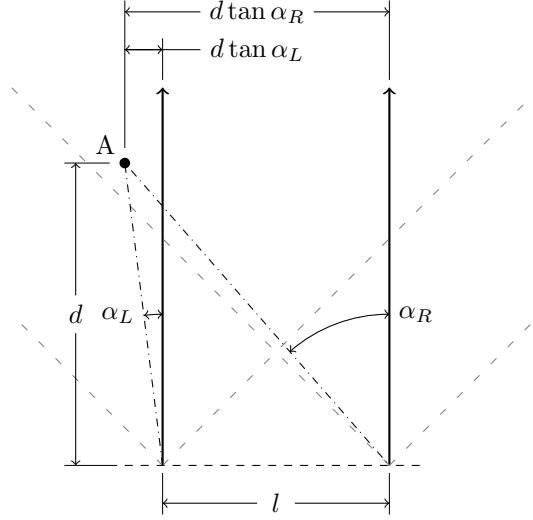$$d = \frac{l}{-\tan \alpha_R + \tan \alpha_L} \tag{4}$$

3

Figure 3: An object in stereoscopic view.

In all three of these equations, we have assumed that α_R and α_L are always positive. However, if we are judicious about how we apply signs to these angles, we can make one equation apply for any placement of Object A. We have three choices. We can assert that

- each angle α is positive for pixels on the right half of the respective image, in which case equation 4 always holds.

- each angle α is positvie for pixels on the left half of the respective image, in which case equation 3 always holds.

- α_R is positive for pixels on the left side of the right image, and α_L is positive for pixels on the right side of the left image, in which case equation 2 always holds.

4   Pixel Correlation

4.1   Cross-Correlation

5   Depth Field Representation

6   Accounting for Barrel Distortion

5